

ADAPTIVE COMMUNICATIONS AND SIGNAL PROCESSING LABORATORY
CORNELL UNIVERSITY, ITHACA, NY 14853

Optimal Placement of Known Symbols for Slowly Varying Frequency Selective Channels

S.Adireddy and L.Tong

Technical Report No. ACSP-TR-04-03-01

April 2003



Abstract

The problem of placing known symbols in a data stream for a slowly varying frequency selective channel is considered from an information-theoretic perspective. Given the amount of redundancy associated with known symbols, placement schemes that minimize the outage probability are derived by assuming that the transmitted codewords consist of packets that are constrained to have the same known symbol placement. Under the assumption that each known symbol cluster is at least as large as $\alpha \geq 2L + 1$ (where L is the channel order), we show that the optimal placement is obtained by arranging the known symbols into as many clusters as possible and placing them such that the unknown symbol blocks are as equal as possible. It is shown that the optimal placement of known symbol clusters does not depend on the probability density of the channel. Under some further constraints, it is shown that the placement schemes derived are optimal in the frame work of error exponents as well. Numerical examples are used to illustrate the ideas and potential gains of using optimal known symbol placement.

Keywords

Known Symbols, Placement Schemes, Compound Channels, Outage Probability, Error Exponents.

I. INTRODUCTION

Although information theory does not mandate the separation of the transmitted signal into known and unknown symbols, training symbols are inserted into the data stream to enable low complexity receiver implementations. For wireless communication, where the channel can vary significantly over time, a large portion of the resources might be dedicated to training symbols. It is therefore important for a system designer to optimize the design of training. This includes optimization of parameters such as percentage of training symbols and placement of training symbols.

In this paper, we consider the problem of optimal placement of training for a slowly varying frequency selective channel. We are particularly interested in the case when the decoding delay of the data application is such that it is not possible to code over multiple fades. The information theoretic performance of such a system is usually analyzed under a composite channel model, which is a compound channel with apriori probabilities on the channel states [1], [2]. This model, first analyzed in [3], has gathered a lot of attention. See for example [4], [5], [6], [7], [8], [9]. In this model, it is assumed that the whole codeword sees a single fade irrespective of its length. The *practical* justification of this model is that the decoding delay is such that the codeword cannot be longer than the coherence time of the channel and the coherence time is long enough for information theoretic results to be meaningful. How long is long enough depends on the symbol error rate required and error exponents of this channel model. See [2] for a good discussion about these issues. Probability of outage is an important measure of performance for this channel model. We will therefore use outage probability as the performance metric to compare different placement schemes.

We assume that the receiver forms an estimate of the channel based on training alone and the estimate is then used by the decoder in order to perform decoding. It is assumed that the number of known

symbols inserted in a codeword is sufficient to form a *reliable* estimate of the channel.[†] This assumption also allows us to utilize the techniques available on outage probability for known channels to address the problem of training symbol placement. The drawback of this assumption is that it does not allow us to determine the optimal percentage of known symbols. The optimal percentage of known symbols can be however be obtained by analyzing the problem in the framework of error exponents because it allows us to take into account the error due to channel estimate. Problem formulation in terms of error exponents is also explored in the paper.

We assume that the codebooks used to transmit the bits consist of *packetized codewords* (PCW). Packetized codewords are constrained to consist of an integral number of *packets*. Each packet is assumed to have the same placement scheme and we are interested in optimizing this placement strategy. This model turns out to be simple to analyze since the restriction of packetization converts the model into a vector discrete memoryless channel, once we view each packet as a symbol. The optimal placement schemes for known symbols within the packet are then obtained with ease. The problem of optimal placement for general codewords where there is no constraint on the placement of known symbols is considered in [10]. For the case of general codewords, since we remove the constraint of packetization, outage probability for the general codeword model should be defined more carefully.

For the PCW case, we show that under the constraint that known symbols are placed in clusters of length at least $\alpha \geq 2L + 1$ where L is the order of the channel, the outage probability is minimized by the family of placement schemes referred to as QPP- α . In this family, the known symbols are broken into as many clusters as possible under the constraint that each of them is at least α and they are placed such that the lengths of unknown symbol clusters are as equal as possible. It turns out that the conventional strategy of placing known symbols together in big clusters carries a penalty in performance. Surprisingly, the optimal known symbol placement does not depend on the probability density of the channel.

We formulate the problem of optimal placement in terms of the error exponents of the channel. We show that for $\alpha \geq 2L + 1$, QPP- α placement schemes are optimal with respect to this metric as well. For this formulation, unlike in the formulation for outage probability, we do not have any restrictions on the structure of code words. The advantage of the error exponent formulation is that it allows one to take into account the error in the channel estimate as well.

There has been some prior work reported on the effect of training and channel estimation errors on mutual information that does not consider the placement issue [11], [12], [13]. Medard [11] has obtained lower and upper bounds on mutual information that are a function of the variance of the error in the

[†]Technically, since the channel stays constant over the duration of the code word and the code word can be as large as possible, it should be possible to obtain an estimate of the channel to the desired accuracy by placing an arbitrarily small percentage of known symbols. The importance of placement can then be questioned. In reality, however, the coherence time of the channel places a limit on the length of the codeword and thus in order to obtain a *reliable* estimate of the channel, it might be necessary to dedicate a significant percentage of the time to transmitting known symbols.

channel estimate formed at the receiver. Hassibi and Hochwald [12] have optimized training in multiple-antenna systems with quasi-static flat fading by maximizing a tight lower bound on the ergodic capacity. They considered issues such as amount of training, choice of training symbols and power allocation. For the channel model considered in their work, the performance is independent of placement. The training issues for quasi-static frequency selective fading were addressed in [13]. In this paper, the placement issue was not considered and the known symbols were placed at the beginning of the packet.

We have addressed the problem of optimal placement of known symbols for ergodic block frequency selective fading with i.i.d Gaussian taps in [14]. We use the framework derived in [12], [13] to obtain a tight lower bound on the channel capacity and maximize this lower bound to obtain optimal placement schemes for both OFDM and single carrier systems. We show that in OFDM placing known symbols periodically in frequency is optimal where as periodic placement in time (QPP- α placement schemes) turns out to be optimal for single carrier systems. We have also previously considered the problem of joint optimization of known symbol placement and equalizer for a symbol-by-symbol decision feedback receiver [15]. The performance criterion used was Average Mean Square Error (AMSE). It turns out that the optimal symbol placement is to separate the known symbols by at least d , the detection delay of the decision feedback receiver. It was shown that substantial gain could be achieved by spreading the known symbols in the data stream instead of clustering them.

For the quasi-static block fading model with independent fades, tracking of the channel is not an issue. Training issues for tracking in Gauss-Markov channels was considered in [16], [17], [18]. In [16] Medard *et al.* optimized the spacing between consecutive pilot symbols by maximizing the capacity with binary signaling. In [17], optimal placement schemes for finite packet transmission over Gauss-Markov channels was considered. Placement schemes for continuous transmission over Gauss-Markov channel were studied by Dong *et al.* in [19].

The optimal placement problem has been explored for other metrics and models too. Optimal placement of known symbols for minimizing the variance of the error in channel estimate for OFDM systems has been addressed in [20]. The optimal placement for the more general setting of block precoded transmission systems with cyclic prefix was addressed with the channel estimate as the metric in [21] and at high SNR, with block length going to infinity, and with ergodic capacity as the metric in [22]. The channel was assumed to be Gaussian with independent but not identical taps and was assumed to under go quasi-static frequency selective fading. All these papers showed that periodic placement in frequency is optimal. Placement issues for multiple-antenna systems employing orthogonal space-time codes with the CRLB for channel estimation as the metric has been considered in [23]. The placement that minimizes the Cramer-Rao Lower Bound (CRLB) for semi-blind channel estimators was found in [24]. It was shown here that QPP- α placement is optimal under some constraints. It is quite surprising that the QPP- α placement schemes turn out to be optimal for a variety of metrics.

This paper is organized as follows. In Section II, we introduce the channel model, the model for

the codebook, the receiver structure and training structure and define outage probability. In Section III we formulate the problem of optimal known symbol placement. In Section IV, we obtain optimal placement schemes. In Section V, we explore the problem formulation of optimal placement in terms of error exponents for the composite channel. Section VI illustrates the ideas proposed in the paper through different simulations. We finally conclude in VII. The appendix contains proofs of some of the results stated in the paper.

II. SYSTEM MODEL

In this section we first define the channel model. We then describe in detail the PCW model. The training schemes used are then described. We also give the structure of the receiver employed by each user. We then introduce the metric that is used for optimizing placement in each case.

A. Channel Model

We assume that the channel $\mathbf{h} = [h_0, h_1, \dots, h_L]^t$ to a user is random and is governed by the density function $p_{\mathbf{h}}(\cdot)$. We also assume that the channel stays constant for over the duration of the code word. We assume that neither the receiver nor the transmitter knows the propagation coefficients. The channel output to each user is corrupted by the additive, zero mean, white Gaussian noise w_k with variance 1. We assume that the average energy of the unknown symbols is equal to ρ .

B. Packetized Codewords

We assume that the codeword consists of packets that belong to the class \mathcal{P}_α . A packet is in the class \mathcal{P}_α if

A1: The length of each packet is $(N + P + L)$ where N is the number of unknown symbols and $(P + L)$ is the number of known symbols.

A2: The known symbols come in clusters of length equal to at least $\alpha \geq 2L + 1$.

A3: Each packet starts with at least L known symbols.

The assumption A2 is introduced primarily from the point of view of channel estimation. A2 makes it possible to employ channel estimators based on training only. A3 implies that there is no inter-packet interference.

As shown in Fig. 2, every placement in the packet can be specified by $\mathbf{r} = (\mathbf{m}, \mathbf{n})$, that is, two tuples $\mathbf{m} = (m_1, \dots, m_J)$ and $\mathbf{n} = (n_1, \dots, n_{J+1})$ where \mathbf{m} gives the lengths of unknown symbol blocks and \mathbf{n} the lengths of known symbol clusters. For placements that end with unknown symbols we have $n_{J+1} = 0$. Further A3 implies that $n_1 \geq L$. We also have $\sum_{i=1}^J m_i = N$ and $\sum_{i=1}^{J+1} n_i = (P + L)$. It should be noted that the number of elements in each of these tuples is a variable and depends on the placement scheme. We refer to the symbols between any two consecutive known symbol clusters as unknown symbol blocks.

A code of rate R whose codewords consist of k packets each is denoted as (k, R) . A rate R is said to be achievable with a placement \mathbf{r} if there exists a sequence of codes $\{(k, R)\}, k = 1, 2, \dots$ such that the

placement of known symbols in each packet is \mathbf{r} and the probability of error tends to zero.

C. Training Structure

For simplicity, we constrain the training symbol clusters to a specific structure. We assume that each training cluster contains only one non-zero training symbol. This symbol can be placed at any position in the cluster that is at least L away from either edge. Since we restrict ourselves to the case $\alpha \geq 2L + 1$, this is always possible. A sample training cluster is shown in Fig. 3. This structure on training plays two roles. Firstly, channel estimation becomes simple to implement and analyze. The structure also ensures that the training is *orthogonal*. It is well known that *orthogonal* training has many desirable properties. Secondly, separating the training symbols into multiple clusters does not come with a penalty in performance. If equal energy training symbols are used, then splitting a training cluster into multiple cluster carries a penalty in performance (assuming training only estimators are employed). The use of training clusters with this structure allows us to decouple the problem of channel estimation from that of known symbol placement to a certain extent.

D. Receiver Structure and Performance

We assume that the transmitter uses a codebook $\mathcal{C} = \{\mathbf{s}^{(1)}, \mathbf{s}^{(2)}, \dots, \mathbf{s}^{(M)}\}$ to transmit the data, and every codeword contains known symbols that enable the receiver to estimate the channel. Fig. 4 illustrates the receiver structure, the channel estimator is given by $\hat{\mathbf{h}} = g(\mathbf{y}, \mathbf{s}_t)$ where \mathbf{y} is the vector containing the received output due to the codeword and \mathbf{s}_t is the vector containing the non-zero known symbols. The codeword is decoded as

$$\hat{\mathbf{s}} = \arg \min_{\mathbf{s} \in \mathcal{C}} p(\mathbf{y} | \mathbf{s}, \hat{\mathbf{h}}), \quad (1)$$

where $p(\mathbf{y} | \mathbf{s}, \hat{\mathbf{h}})$ is the conditional probability density of \mathbf{y} conditioned on the transmitted code word and the estimated channel.

E. Outage Probability as Performance Metric

Due to the constraint that each code word undergoes a single fade, it is possible that the Shannon capacity of this channel is equal to zero. A meaningful metric for this channel is therefore outage probability. The interpretation of outage probability is the one given in [6] and the references therein. Given a transmitted SNR ρ and rate R , $\Theta(R, \rho)$ is the largest possible set for which C_Θ , the capacity of the compound channel with the parameter $\mathbf{h} \in \Theta(R, \rho)$, satisfies $C_\Theta \geq R$. The outage probability is then defined as $P_{out}(R, \rho) = \Pr(\mathbf{h} \notin \Theta(R, \rho))$. Hence there exists a sequence of codes \mathcal{C}_n of rate R that satisfies the power requirement ρ , and for which the supremum of the probability of error over all channels $\mathbf{h} \in \Theta(R, \rho)$ tends to zero.

For problem formulation in terms of outage probability, we assume that number of training symbols inserted is sufficient to form a reliable estimate. This allows us to use the literature for known channels

to obtain the expression for outage probability. In this paper, we consider outage probability only for i.i.d inputs. This makes the problem tractable and it also implies that the capacity achieving codes will have a flat transmit spectrum.

III. PROBLEM FORMULATION

The use of packetized codewords turns the model into the vector discrete memoryless channel shown in Fig.5. Since the first L symbols of a packet are known, the contribution due to these symbols can be subtracted without any loss of generality. The column vector $\mathbf{s} = [s_1, \dots, s_{N+P}]^t$ contains the symbols in the packet excluding the first $(L+1)$ and \mathbf{y} is the output obtained after subtracting the contribution due to the starting $(L+1)$ known symbols. The relation between \mathbf{s} and \mathbf{y} is given by

$$\mathbf{y} = \mathbf{H}_{N+P} \mathbf{s} + \mathbf{w}, \quad (2)$$

where

$$\mathbf{H}_{N+P} = \begin{bmatrix} h_0 & 0 & \cdots & 0 \\ h_1 & h_0 & & \vdots \\ \vdots & h_1 & & 0 \\ h_L & \vdots & & h_0 \\ 0 & h_L & & h_1 \\ \vdots & 0 & & \vdots \\ 0 & \cdots & \cdots & h_L \end{bmatrix}_{(N+P+L) \times (N+P)}. \quad (3)$$

The probability of outage for this model is a function of the placement scheme \mathbf{r} in the packet. The expression for $P_{out}(R, \rho, \mathbf{r})$ is given by [5]

$$P_{out}(R, \rho, \mathbf{r}) = P(\Psi(\mathbf{r}, \rho, \mathbf{h}) < R), \quad (4)$$

where

$$\Psi(\mathbf{r}, \rho, \mathbf{h}) = \log \det(\mathbf{I} + \rho \mathbf{H}_{N+P} \mathbf{J}_{\mathbf{r}} \mathbf{H}_{N+P}^H). \quad (5)$$

The matrix $\mathbf{J}_{\mathbf{r}}$ is a square diagonal selection matrix of order $(N+P)$. It has ones in those positions that have unknown symbols and zeros in those that have known symbols. If $\mathbf{r} = (\mathbf{m}, \mathbf{n})$, then

$$\Psi(\mathbf{r}, \rho, \mathbf{h}) = \log \prod_{i=1}^J \det(\mathbf{I} + \rho \mathbf{H}_{m_i} \mathbf{H}_{m_i}^H) \quad (6)$$

$$\triangleq \log \prod_{i=1}^J G(m_i, \rho, \mathbf{h}) \quad (7)$$

This is possible due to A3 (each non-zero element of \mathbf{n} is at least as large as L , which implies that there is no inter-symbol interference between two unknown symbol blocks). We see that the outage probability depends only on \mathbf{m} . The placements corresponding to all the permutations of the elements of the tuple

\mathbf{m} have the same outage probability. We can thus conclude that the order in which the unknown symbol blocks or the known symbol clusters are transmitted is immaterial. In fact the exact values of \mathbf{n} are also irrelevant as long as A1 is satisfied and the channel estimator is consistent.

The objective is to minimize the outage probability with respect to the size of unknown symbol blocks \mathbf{m} . Formally, our objective is to examine the following optimization :

$$\mathbf{m}^* = \arg \min_{\mathbf{m} \in \mathcal{P}_\alpha} P_{out}(R, \rho, \mathbf{r}). \quad (8)$$

IV. QUASI PERIODIC PLACEMENT AND ITS OPTIMALITY

In this section, we define a family of placement schemes for the PCW model called Quasi Periodic Placement (QPP) and prove that they are optimal. We also give some alternate placement schemes that have good performance.

The family of QPP placement schemes is divided into different classes on the basis of the smallest allowable length of any known symbol cluster. The class of schemes for which α is the smallest allowable known symbol cluster length is denoted as QPP- α . Formally, we define \mathcal{Q}_α as the set of all placement schemes belonging to the class QPP- α .

Definition 1: Given an α and a packet with N unknown symbols and $(P + L)$ known symbols, let $J = \lfloor \frac{P+L}{\alpha} \rfloor$. A placement scheme belongs to \mathcal{Q}_α if and only if

1. $\mathbf{n} \in \mathcal{N}$ where $\mathcal{N} = \{(n_1, \dots, n_J, 0) : \sum_i n_i = (P + L), \min(\{n_1, \dots, n_J\}) \geq \alpha\}$
2. $\mathbf{m} \in \mathcal{M}$ where $\mathcal{M} = \{(m_1, \dots, m_J) : \sum_i m_i = N, m_i \in \{\lfloor \frac{N}{J} \rfloor, (\lfloor \frac{N}{J} \rfloor + 1)\}\}$

In other words, in a QPP- α placement scheme, the known symbols are divided into as many clusters as possible under the constraint that each of them is no less than α , and these clusters are placed such that the unknown symbol blocks are as “equal” as possible. An element in \mathcal{Q}_α is denoted as \mathbf{r}^α .

We state some fairly straight forward properties of QPP- α schemes without proof.

P1: All QPP schemes have at most two different unknown symbol block sizes. If there are two distinct unknown symbol block sizes, they differ by one.

P2: Under assumptions A1 and A2, given $\alpha \geq 2L + 1$, all the placement schemes in \mathcal{Q}_α have the same outage probability. For all other α , such a claim is not true in general.

We now state a property of the function $G(n, \rho, \mathbf{h})$, defined in (7), that will be useful in proving the optimality of QPP- α schemes.

Lemma 1: The function $G(n, \rho, \mathbf{h})$, defined in (7), has following property :

$$G(n, \rho, \mathbf{h})^2 \geq G(n + k, \rho, \mathbf{h})G(n - k, \rho, \mathbf{h}) \quad \forall n \in \mathcal{Z}^+, k \in \{0, 1, \dots, n\}. \quad (9)$$

Proof: Refer to the Appendix.

We now state and prove the optimality properties of the QPP- α schemes. For any ρ and \mathbf{h} , placement schemes in QPP- α maximize $\Psi(\mathbf{r}, \rho, \mathbf{h})$.

Lemma 2: If $\mathbf{r}^\alpha \in \mathcal{Q}_\alpha$ then under A1-A3

$$\Psi(\mathbf{r}^\alpha, \rho, \mathbf{h}) = \max_{\mathbf{m} \in \mathcal{P}_\alpha} \Psi(\mathbf{r}, \rho, \mathbf{h}) \quad (10)$$

$$= \max_{\mathbf{m} \in \mathcal{P}} \log \prod_{i=1}^J G(m_i, \rho, \mathbf{h}) \quad (11)$$

$$= (J^* - n) \log b_{N_1} + n \log b_{N_1+1}, \quad (12)$$

where $J^* = \lfloor \frac{P+L}{\alpha} \rfloor$, $N_1 = \lfloor \frac{N}{J^*} \rfloor$, $n = N - J^* N_1$, and $b_n \triangleq |\mathbf{I}_n + \rho \mathbf{H}_n^H \mathbf{H}_n|$.

Proof : Using the property of the function $G(n, \rho, \mathbf{h})$ stated in Lemma 1, we can say that for any given \mathbf{r} , $\Psi(\mathbf{r}, \rho, \mathbf{h})$ reaches a maximum when J is as large as possible, and every element of the tuple m_i is as small as possible. Therefore we must have $J = J^* = \lfloor \frac{P+L}{\alpha} \rfloor$ and $\mathbf{m} \in \mathcal{M}$ where

$$\mathcal{M} = \{(m_1, \dots, m_J) : \sum_i m_i = N, m_i \in \{\lfloor \frac{N}{J^*} \rfloor, (\lfloor \frac{N}{J^*} \rfloor + 1)\}\}. \quad (13)$$

All the placement schemes belonging to the class QPP- α satisfy this criterion.

From the property P1, we have that there are at most two distinct unknown symbol block sizes. Let the length of the smaller unknown symbol block be N_1 . Let the number of blocks with length $(N_1 + 1)$ be equal to n . Then we can easily show that $N_1 = \lfloor \frac{N}{J^*} \rfloor$ and $n = N - N_1 J^*$. From (7), it can be seen that

$$\begin{aligned} \Psi(\mathbf{r}^\alpha, \rho, \mathbf{h}) &= (J^* - n) \log G(N_1, \rho, \mathbf{h}) + n \log G(N_1 + 1, \rho, \mathbf{h}) \\ &= (J^* - n) \log b_{N_1} + n \log b_{N_1+1}. \end{aligned}$$

□

Using Lemma 2 with the definition of outage probability in (4), we can show that all placement schemes in QPP- α minimize the outage probability.

Theorem 1: If $\mathbf{r}^\alpha \in \mathcal{Q}_\alpha$, then under A1-A3, for any given $p_{\mathbf{h}}(\cdot)$, R and ρ

$$P_{out}(R, \rho, \mathbf{r}^\alpha) = \min_{\mathbf{r} \in \mathcal{P}_\alpha} P_{out}(R, \rho, \mathbf{r}). \quad (14)$$

Furthermore, $P_{out}(R, \rho, \mathbf{r}^\alpha)$ is a monotonically increasing function of α . Hence

$$P_{out}(R, \rho, \mathbf{r}^L) = \min_{\alpha \geq L, \mathbf{r} \in \mathcal{P}_\alpha} P_{out}(R, \rho, \mathbf{r}). \quad (15)$$

Theorem 1 shows that if we allow all possible $\alpha \geq 2L + 1$, the placement schemes belonging to QPP- $2L + 1$ are optimal. Conventionally known symbols have been placed in big clusters. Theorem 1 indicates that there is some gain to be achieved by spreading them. The algorithm for placing the known symbols is also quite simple. The optimal placement is independent of the probability density of the channel coefficients. This property makes the scheme highly attractive for the broadcast scenario. The intuition in placing the known symbols in small clusters is that known symbol clusters reduce the inter-symbol interference (ISI) and one should thus maximize the number of known symbol clusters in the data stream.

A. An Example

We explain the algorithm further with an example. Consider a transmission system for which $N = 116, P = 29$, and $L = 2$. If $\alpha = 5$, then the optimal placement scheme belong the class QPP-5. Then, we have total number of unknown symbol blocks

$$J^* = \lfloor \frac{P+L}{2L+1} \rfloor = 6. \quad (16)$$

The minimum number of unknown symbols in each block is equal to $\lfloor \frac{N}{J^*} \rfloor = 19$. So unknown symbols come in clusters of length equal to either 19 or 20. There are 4 blocks of length 19 and 2 blocks of length 20. The order in which these unknown symbol blocks are sent is immaterial.

V. ERROR EXPONENTS

In this section, we will formulate the problem of optimal placement in terms of error exponents. This formulation has the advantage that it is possible to take into account the error in channel estimate. In addition to optimal placement of training, this formulation can be potentially used to determine the optimal percentage of known symbols. However, in this paper, we will focus on obtaining a simple upper bound on the average probability of error of a random code book. We will then optimize the placement of known symbols based on the upper bound.

Consider a code book such that each code word is of length T , the number of unknown symbols is equal to N and the number of known symbols is equal to P . Further we assume that each codeword start with L zeros and ends with L zeros. Thus $T = N + P + 2L$. We assume that $T \geq 2L + 1$ and $P \geq 2L + 1$ and that orthogonal training is employed. It is assumed that the total energy in the training symbols is given by $P\rho$ and the data symbols is given by $N\rho$. Let the placement of known symbols be given by the tuple \mathbf{r}^\dagger . For the purposes of this section, we will assume that the channel coefficients are i.i.d circularly complex Gaussian with mean 0 and variance equal to $\frac{1}{L+1}$. We assume that the channel estimator forms an MMSE estimate of the channel and the decoder performs ML-decoding based on the channel estimate. If ρ is the average energy of the data and training we will show that $P_e(\rho, R, N, P, L, \mathbf{r})$, an upper bound on the average probability of error is given by

$$P_e(\rho, R, N, P, L, \mathbf{r}) \leq \exp(-TE_r(\rho, R, N, P, L, \mathbf{r})), \quad (17)$$

where

$$E_r(\rho, R, N, P, L, \mathbf{r}) = \max_{t \in (0,1)} E_0(t, \rho, N, P, L, \mathbf{r}) - tR, \quad (18)$$

and

$$E_0(t, \rho, N, P, L, \mathbf{r}) = -\log E \left\{ \prod_{i=1}^J \det \left(\mathbf{I} + \frac{\rho}{(1+t)} \frac{P\rho}{P\rho + (L+1)(\rho+1)} \mathbf{H}_{m_i}^H \mathbf{H}_{m_i} \right)^{-t} \right\}$$

[†]Note that in this formulation, there is no restriction on the placement of known symbols with in the codeword

$$= -\frac{1}{T} \log \mathbb{E} \left\{ \prod_{i=1}^J G(m_i, \rho_t, \mathbf{h}) \right\}^{-t}, \quad (19)$$

where

$$\rho_t = \frac{\rho}{(1+t)} \frac{P\rho}{(L+1)(\rho+1) + P\rho}. \quad (20)$$

In what follows we will show that QPP- α placement is optimal based on $P_e(\rho, R, N, P, L, \mathbf{r})$. That is :

Theorem 2: If \mathbf{r}^α refers to a QPP- α placement scheme and $\alpha \geq (2L+1)$,

$$E_r(\rho, R, N, P, L, \mathbf{r}^\alpha) = \max_{\mathbf{r} \in \mathcal{P}^\alpha} E_r(\rho, R, N, P, L, \mathbf{r}). \quad (21)$$

That is,

$$P_e(\rho, R, N, P, L, \mathbf{r}^\alpha) = \min_{\mathbf{r} \in \mathcal{P}^\alpha} P_e(\rho, R, N, P, L, \mathbf{r}) \quad (22)$$

A. Channel Estimation

The channel model assumed and the structure of the training allows us to estimate each tap separately. The model for estimation is given by

$$\mathbf{y}_t^{(i)} = \mathbf{s}_t h_i + \mathbf{n}_t, \quad (23)$$

where \mathbf{s}_t is a column vector containing all the non-zero training symbols and h_i is the i^{th} coefficient of the channel. The MMSE estimate \hat{h}_i is equal to

$$\hat{h}_i = \frac{1}{(L+1) + \mathbf{s}_t^H \mathbf{s}_t} \mathbf{s}_t^H \mathbf{y}_t^{(i)}. \quad (24)$$

The mean square error of the estimate is given by

$$\mathbb{E}\{\tilde{h}_i^2\} = \frac{1}{\mathbf{s}_t^H \mathbf{s}_t + (L+1)}. \quad (25)$$

It is easy to see that the auto-correlation of $\tilde{\mathbf{h}}$ which is the error in the channel estimate is diagonal and is given by

$$\mathbf{R}_e = \mathbb{E}\{\tilde{h}_i^2\} \mathbf{I}. \quad (26)$$

B. Data Model

The relationship between the output due to unknown symbols \mathbf{y}_d and unknown symbols \mathbf{s}_d is given by

$$\underbrace{\begin{bmatrix} \mathbf{y}_{1d} \\ \mathbf{y}_{2d} \\ \vdots \\ \mathbf{y}_{Jd} \end{bmatrix}}_{\mathbf{y}_d} = \underbrace{\begin{bmatrix} \mathbf{H}_{m_1} & 0 & \cdots & 0 \\ 0 & \mathbf{H}_{m_2} & & 0 \\ \vdots & & \ddots & \\ 0 & & & \mathbf{H}_{m_J} \end{bmatrix}}_{\mathbf{H}_d} \underbrace{\begin{bmatrix} \mathbf{s}_{1d} \\ \mathbf{s}_{2d} \\ \vdots \\ \mathbf{s}_{Jd} \end{bmatrix}}_{\mathbf{s}_d} + \mathbf{n}_d. \quad (27)$$

The matrix \mathbf{H}_{m_i} is a toeplitz matrix of size $(m_i + L) \times m_i$ given by

$$\mathbf{H}_{m_i} = \begin{bmatrix} h_0 & 0 & \cdots & 0 \\ h_1 & h_0 & & \vdots \\ \vdots & h_1 & & 0 \\ h_L & \vdots & & h_0 \\ 0 & h_L & & h_1 \\ \vdots & 0 & & \vdots \\ 0 & \cdots & \cdots & h_L \end{bmatrix}_{(m_i+L) \times (m_i)}. \quad (28)$$

The matrix \mathbf{H}_d has this structure because of the structure imposed on the training clusters. The input output relation can be further written as

$$\mathbf{y}_d = \hat{\mathbf{H}}_d \mathbf{s}_d + \underbrace{\tilde{\mathbf{H}}_d \mathbf{s}_d + \mathbf{n}_d}_{\mathbf{v}_d}. \quad (29)$$

The matrix $\hat{\mathbf{H}}_d$ refers to the matrix formed by the estimate of the channel. The decoder has access to both \mathbf{y}_d and $\hat{\mathbf{H}}_d$. We assume that ML decoding is performed. That is,

$$\hat{\mathbf{s}}_d = \arg \max_{\mathbf{s}_d \in \mathcal{C}} p(\mathbf{y}_d | \mathbf{s}_d, \hat{\mathbf{H}}_d). \quad (30)$$

C. Upper Bound on Probability of Error

An upper bound on the probability of error of this decoder averaged over codes can be found by the method of error exponents [25], [8], [9]. It is given by

$$P_e(\rho, R, N, P, L, \mathbf{r}) = \exp(-T(E_r(\rho, R, N, P, L, \mathbf{r}))), \quad (31)$$

where

$$E_r(\rho, R, N, P, L, \mathbf{r}) = \max_{t \in (0,1)} E_0(t, \rho, N, P, L, \mathbf{r}) - tR \quad (32)$$

and

$$E_0(t, \rho, N, P, L, \mathbf{r}) = -\frac{1}{T} \log \mathbb{E} \left\{ \int_{\mathbf{y}_d} \left(\int_{\mathbf{s}_d} p(\mathbf{s}_d) p(\mathbf{y}_d | \mathbf{s}_d, \hat{\mathbf{H}}_d)^{\frac{1}{1+t}} d\mathbf{s}_d \right)^{(1+t)} d\mathbf{y}_d \right\}. \quad (33)$$

We fix the input distribution be i.i.d Gaussian. However, it is difficult to evaluate $E_0(\cdot)$ analytically because the distribution of \mathbf{v}_d conditioned on \mathbf{s}_d and $\hat{\mathbf{H}}_d$ is Gaussian whose mean and covariance depend on \mathbf{s}_d . Hence, we upper bound the average probability of error by choosing the distribution of \mathbf{v}_d as a *worst case* distribution. We know that for a given $p(\mathbf{s}_d)$, $E_r(\cdots) \geq 0$ if and only if

$$0 \leq R \leq I(\mathbf{s}_d; \mathbf{y} | \hat{\mathbf{H}}_d). \quad (34)$$

We choose the distribution of \mathbf{v}_d to be that distribution which minimizes $I(\mathbf{s}_d; \mathbf{y} | \hat{\mathbf{H}}_d)$. In order to obtain a tight bound, we constrain the set of distributions to be those that retain some first order and second

order properties of $\mathbf{v}_d = \tilde{\mathbf{H}}_d \mathbf{s}_d + \mathbf{w}_d$. It can be seen that

$$\begin{aligned} \mathbb{E} \left\{ \mathbf{v}_d | \hat{\mathbf{H}}_d \right\} &= \mathbf{0} \\ \mathbb{E} \left\{ \mathbf{v}_d \mathbf{v}_d^H | \hat{\mathbf{H}}_d \right\} &= \mathbf{R} + \mathbf{I} \end{aligned} \quad (35)$$

$$\mathbb{E} \left\{ \mathbf{s}_d \mathbf{v}_d^H | \hat{\mathbf{H}}_d \right\} = \mathbf{0}. \quad (36)$$

The last property follows due to the MMSE nature of the estimate. The matrix \mathbf{R} is given by

$$\rho \mathbb{E} \{ \tilde{\mathbf{H}}_d \tilde{\mathbf{H}}_d^H \} = \begin{bmatrix} \rho \mathbb{E} \{ \tilde{\mathbf{H}}_{m_1} \tilde{\mathbf{H}}_{m_1}^H \} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{0} & \rho \mathbb{E} \{ \tilde{\mathbf{H}}_{m_2} \tilde{\mathbf{H}}_{m_2}^H \} & & \mathbf{0} \\ \vdots & & \ddots & \\ \mathbf{0} & & & \rho \mathbb{E} \{ \tilde{\mathbf{H}}_{m_J} \tilde{\mathbf{H}}_{m_J}^H \} \end{bmatrix} \quad (37)$$

Each of the matrix $\mathbb{E} \{ \tilde{\mathbf{H}}_{m_i} \tilde{\mathbf{H}}_{m_i}^H \}$ is a diagonal matrix, since errors in the estimates of the taps are uncorrelated.

Given $p(\mathbf{s}_d)$, we define $\Omega(p(\mathbf{s}_d))$ the set of all the conditional probability distributions that retain the properties of \mathbf{v}_d listed above. That is

$$\begin{aligned} \Omega(p(\mathbf{s}_d)) &= \left\{ p(\mathbf{n}_d | \mathbf{s}_d, \hat{\mathbf{H}}_d) : \mathbb{E} \left\{ \mathbf{n}_d | \hat{\mathbf{H}}_d \right\} = \mathbf{0}, \mathbb{E} \left\{ \mathbf{n}_d \mathbf{n}_d^H | \hat{\mathbf{H}}_d \right\} = \mathbb{E} \left\{ \mathbf{v}_d \mathbf{v}_d^H | \hat{\mathbf{H}}_d \right\}, \right. \\ &\quad \left. \mathbb{E} \left\{ \mathbf{n}_d \mathbf{s}_d^H | \hat{\mathbf{H}}_d \right\} = \mathbb{E} \left\{ \mathbf{v}_d \mathbf{s}_d^H | \hat{\mathbf{H}}_d \right\} = \mathbf{0} \right\}. \end{aligned} \quad (38)$$

We choose the probability distribution of the residual noise as the one in the set $\Omega(p(\mathbf{s}_d))$ which minimizes $I(\mathbf{s}_d; \mathbf{y} | \hat{\mathbf{H}}_d)$. It can be shown that the worst case distribution is independent Gaussian with zero mean and covariance equal to $\mathbf{R}_v \triangleq \mathbf{R} + \mathbf{I}$ [12], [13], [14]. For a detailed proof of this result on the lines of [12], [11], see [10]. For this choice of residual noise, it can be easily shown [5] that

$$\begin{aligned} E_0(t, \rho, N, P, L, \mathbf{r}) &= -\frac{1}{T} \log \mathbb{E} \left\{ \det \left(\mathbf{I} + \frac{\rho}{1+t} \mathbf{R}_v^{-1} \hat{\mathbf{H}}_d \hat{\mathbf{H}}_d^H \right)^{-t} \right\} \\ &\leq -\frac{1}{T} \log \mathbb{E} \left\{ \det \left(\mathbf{I} + \frac{\rho}{(1+t)} \frac{\mathbf{s}_t^H \mathbf{s}_t}{(L+1)(\rho+1) + \mathbf{s}_t^H \mathbf{s}_t} \mathbf{H}_d \mathbf{H}_d^H \right)^{-t} \right\} \\ &= -\frac{1}{T} \log \mathbb{E} \left\{ \det \left(\mathbf{I} + \frac{\rho}{(1+t)} \frac{P\rho}{(L+1)(\rho+1) + P\rho} \mathbf{H}_d^H \mathbf{H}_d \right)^{-t} \right\} \\ &= -\frac{1}{T} \log \mathbb{E} \left\{ \prod_{i=1}^J \det \left(\mathbf{I} + \frac{\rho}{(1+t)} \frac{P\rho}{(L+1)(\rho+1) + P\rho} \mathbf{H}_{m_i}^H \mathbf{H}_{m_i} \right)^{-t} \right\} \\ &= -\frac{1}{T} \log \mathbb{E} \left\{ \prod_{i=1}^J G(m_i, \rho_t, \mathbf{h}) \right\}^{-t}, \end{aligned} \quad (39)$$

where

$$\rho_t = \frac{\rho}{(1+t)} \frac{P\rho}{(L+1)(\rho+1) + P\rho}. \quad (40)$$

The second inequality follows because[§]

$$\mathbf{R}_v \leq \frac{(L+1)}{\mathbf{s}_t^H \mathbf{s}_t + (L+1)} \mathbf{I}. \quad (41)$$

The reason we use this particular inequality is because it is simpler to analyze. Such a technique was first used in [13]. The matrix \mathbf{H}_d is obtained from $\hat{\mathbf{H}}_d$ by scaling each element such that the variance of the channel coefficients is equal to $\frac{1}{L+1}$. The third equality follows because $\mathbf{s}_t^H \mathbf{s}_t = P\rho$.

The problem of optimal placement can now formulated as one that maximizes the quantity $E_r(\rho, R, N, P, L, \mathbf{r})$. In order to perform this maximization, first consider the variation of $E_0(t, \rho, N, P, L, \mathbf{r})$ with placement. Due to the results in Lemma 1 and Lemma 2 and the fact that $E_0(t, \rho, N, P, L, \mathbf{r})$ is an increasing function of

$$\prod_{i=1}^J G(m_i, \rho_t, \mathbf{h}), \quad (42)$$

it easy to see that $E_0(t, \rho, N, P, L, \mathbf{r})$ reaches a maximum only if $\mathbf{r} \in \mathbf{r}^\alpha$ and now since the optimal placement for $E_0(t, \rho, N, P, L, \mathbf{r})$ does not depend on t , we have

$$E_r(\rho, R, N, P, L, \mathbf{r}^\alpha) = \max_{\mathbf{r} \in \mathcal{P}^\alpha} E_r(\rho, R, N, P, L, \mathbf{r}). \quad (43)$$

Theorem 2 then follows.

VI. SIMULATIONS

The composite model that is used in this paper can be used to model a broadcast communication system where the transmitter is transmitting common information to all the receivers. Outage probability in this context refers to the probability that a user is unable to receive the transmission. In our simulations, we will consider the broadcast system example.

A. Ricean Fading Channels

We assume that the receiver might belong to one of three different geographical locations, each of which has a different multi-path structure. Each geographical region is assumed to have line of sight but distinctly different kind of ISI channel. The specular component in region A is assumed to be flat. The specular component in region B is assumed to have nulls on the unit circle where as the specular component in region C is assumed to have a deep null in the spectrum. The goal of the transmitter is to minimize the outage probability for a given rate.

We assume that the channel in each region has $L+1$ taps. The channel in region A is given by

$$h_A(l) = \sqrt{\beta} g_A(l) + \sqrt{1-\beta} r_A(l) \quad l = 0, 1, \dots, L \quad (44)$$

where β gives the power of the specular component. If $g_A(z)$ denotes the z-transform of $g_A(l)$, then we have $g_A(z) = 1$. Hence the specular component is just the delta function. The channel $\{r_A(l)\}_{l=0}^L$ is generated

[§] Given two Hermitian matrices \mathbf{A} and \mathbf{B} , we say $\mathbf{A} \geq \mathbf{B}$ if and only if the matrix $(\mathbf{A} - \mathbf{B})$ is positive semi-definite.

from i.i.d. complex Gaussian with zero mean and variance equal to $\frac{1}{L+1}$. We have analogous equations for channels for region B and region C. For region B, the z-transform of the specular component is given by $g_B(z) = k_1 \prod_{l=0}^{L-1} \left(z^{-1} - \exp(\frac{j2\pi l}{L}) \right)$. The constant k_1 is selected so that the norm of the channel is equal to one. For region C, the z-transform of the specular component is given by $g_C(z) = k_2 (z^{-1} + 1)^L$ where once again the constant k_2 is selected so that the norm of the channel is equal to one.

Fig. 6 compares the performance of the QPP- $2L + 1$ placement scheme with placing all the known symbols at the beginning of the packet (preamble scheme). We assume that $N = 112$, $(P + L) = 48$, $L = 3$ and the transmitted SNR is equal to 20 dB. The user is assumed to belong to one of regions with equal probability. As expected the QPP- $(2L + 1)$ scheme is better than the preamble scheme at every rate. We also find that the gain of the QPP- $(2L + 1)$ scheme is higher at lower outage probabilities. This is because at lower outage probabilities, the bottle neck channels belong to region C, which is where the ISI is greatest. The optimal known symbol placement, being primarily a measure to decrease ISI, provides the maximum gain for these channels.

Fig. 7 plots the variation of the rate achieved with SNR. We find that the gain from optimal placement increases with SNR. This is to be expected since at low SNR, performance degradation is primarily due to noise and at high SNR the performance bottle neck is ISI.

Fig. 8 illustrates the performance of various placement schemes for $L = 1$ scenario. Here we assume that the user can belong to either region A or region C with equal probability.

Fig. 9 shows the variation of the outage probability with SNR for the rate $R = 3.5$ bps per hertz and $L = 1$. We see that it is possible to obtain large gains in outage probability using the optimal placement scheme. This has a direct implication on the coverage that can be obtained at a given SNR.

B. Rayleigh Fading Channels

In this section, we compare the performance of the various placement schemes for the packet codeword scenario for the Rayleigh fading model. We assume that the channel of each user is of length $(L + 1)$. Each tap of the channel is chosen i.i.d complex Gaussian with mean zero and variance $\frac{1}{L+1}$. Fig. 10 shows the performance for this scenario. We see that even though the QPP- $(2L + 1)$ scheme is optimal here, the gain achieved is quite small. This is because for this channel model, unlike the previous channel model, the performance degradation due to ISI is quite small. This example illustrates that the gain achieved from the optimal placement scheme depends on the channel model used.

VII. CONCLUSIONS

In this paper we studied the optimization of placement of known symbols in the data stream for a slowly varying frequency selective channel. The performance metric used is outage probability with i.i.d inputs. We examined the optimization of the position of known symbols in the data stream under the assumption that each codeword consists of packets and each packet contains the same number of known symbols and

the same placement. We show that under the constraint each known symbol cluster is at least of length $\alpha > (2L + 1)$ the outage probability is minimized by breaking the known symbols into as many clusters as possible and by doing the best for placing these clusters periodically in the data stream. In particular the placement schemes belonging to the class QPP- α are optimal.

In [10], we have also addressed the problem of optimal known symbol placement removing the constraint of packetization. We have shown that given the percentage of known symbols $\frac{u}{v}$, the optimal placement schemes are those that belong to the class RQPP- α . These are periodic placement schemes in which the percentage of known symbols is equal to $\frac{u}{v}$ in each period. Further each known symbol cluster is of length exactly α . The known symbol clusters are placed such that the unknown symbol block lengths are as equal as possible.

We also formulated the problem of placement in terms of error exponents and showed that under some further constraints QPP- α schemes remain to be optimal.

Simulations indicate that there is gain to be obtained by optimizing the position of known symbols. The gain in optimization increases with SNR since the optimizing the placement is primarily a tool to decrease ISI. For the same reason, we see that the gain is higher in channel ensembles that are more severely affected by ISI. The placement schemes are optimal given any probability density $p_{\mathbf{h}}(\cdot)$ governing the channel realization.

APPENDIX

I. PROOF OF LEMMA 1

We have

$$G(n, \rho, \mathbf{h}) = |\mathbf{I}_n + \rho \mathbf{H}_n^H \mathbf{H}_n| \quad (45)$$

$$= |\mathbf{I}_n + \rho \mathbf{H}_n^H \mathbf{H}_n| \quad (46)$$

$$\triangleq |D_n|. \quad (47)$$

We have used the fact that $|\mathbf{I} + \mathbf{BC}| = |\mathbf{I} + \mathbf{CB}|$ in deriving the above equations [26]. D_n is a toeplitz matrix of size $n \times n$. The matrix D_n is the n^{th} order principal sub-matrix of the toeplitz matrix D_N . It is known that if K is a positive definite toeplitz matrix and K_n denotes the n^{th} order principal sub matrix of K , then $\frac{|K_n|}{|K_{n-1}|}$ is decreasing in n [27]. From this it follows that

$$|D_n|^2 \geq |D_{n+1}| |D_{n-1}| \quad (48)$$

This proves the lemma for $k = 1$. The result can be proved formally for all k using mathematical induction. \square

REFERENCES

- [1] M.Effros and A.Goldsmith, "Capacity definitions and coding strategies for general channels with receiver side information," in *Proceedings of ISIT*, (Cambridge, MA.), pp. 39-39, August 1998.

- [2] R. Berry and R. Gallager, "Communication over fading channels with delay constraints," *IEEE Transactions on Information Theory*, vol. 48, pp. 1135–1149, May 2002.
- [3] S. L.H.Ozarow and A.D.Wyner, "Information Theoretic Considerations for Cellular Mobile Radio," *IEEE Trans. Vehicular Technology*, vol. 43, pp. 359–378, May 1994.
- [4] G. Foschini and M. Gans, "On limits of wireless communications in a fading environment when using multiple antennas," *Wireless Personal Communications*, vol. 6, no. 3, pp. 311–335, 1996.
- [5] I. Telatar, "Capacity of Multi-antenna Gaussian Channels," *European Trans. Telecomm*, vol. 10, pp. 585–596, Nov-Dec 1999.
- [6] E. Biglieri, J. Proakis, and S. Shamai, "Fading channels: information-theoretic and communications aspects," *IEEE Trans. Inform. Theory*, vol. 44, Oct 1998.
- [7] G.Caire, G.Taricco, and E.Biglieri, "Optimum Power Control over Fading Channels", *IEEE Trans. Info. Theory*, vol. 45, pp. 1468–1489, July 1999.
- [8] G.Kaplan and S.Shamai, "Error Probabilities for the Block-Fading Gaussian Channel," *Archiv fur Elektronik und Uebertragungstechnik*, vol. 47, pp. 192–205, July 1993.
- [9] G.Kaplan and S.Shamai, "Information Rates and Error Exponents of Compound Channels with Application to Antipodal Signaling in a Fading Environment," *Archiv fur Elektronik und Uebertragungstechnik*, vol. 49, pp. 228–39, July 1995.
- [10] S.Adireddy, *The Use of Channel State in Wireless Communication*. PhD thesis, Cornell University, Ithaca, NY, May 2003.
- [11] M. Medard, "The effect upon channel capacity in wireless communication of perfect and imperfect knowledge of the channel," *IEEE Trans. Information Theory*, vol. 46, pp. 933–946, May 2000.
- [12] B. Hassibi and B. Hochwald, "How much training is needed in multiple-antenna wireless links," *Submitted to IEEE Trans. Information Theory*, October 2001.
- [13] H. Vikalo, B. Hassibi, B. Hochwald, and T. Kailath, "Optimal training for frequency-selective fading channels," in *Proc. of Intl. Conf. on ASSP*, (Salt Lake City, Utah), pp. 2105–2108, May 2001.
- [14] S. Adireddy, L. Tong, and H. Viswanathan, "Optimal placement of known symbols for frequency-selective block-fading channels," *IEEE Trans. Info. Theory*, vol. 48, pp. 2338–2353, August 2002. <http://www.ece.cornell.edu/~ltong>.
- [15] S. Adireddy and L. Tong, "Detection with embedded known symbols : optimal symbol placement and equalization ,", in *Proc. ICASSP'00*, vol. 5, (Istanbul, Turkey), pp. 2541–2544, June 2000.
- [16] M. Medard, I. Abou-Faycal, and U. Madhow, "Adaptive coding with pilot signals ,", in *Proc. 38th Annual Allerton Conf. on Communication, Control and Computing.*, (Allerton, IL), October 2000.
- [17] M. Dong, S.Adireddy, and L. Tong, "Optimal Pilot Placement for Semi-Blind Channel Tracking of Packetized Transmission over Time-Varying Channels," *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, vol. E86-A, pp. 550–563, March 2003. <http://acsp.ece.cornell.edu/pubJ.html>.
- [18] S. Adireddy and L. Tong, "Optimal training for time varying ISI channels," in *In Proceedings ISIT'02*, vol. 1, (Lausanne, Switzerland), pp. 271–271, June 2002.
- [19] M. Dong, L. Tong, and B. Sadler, "Optimal Placement of Training over Time-Varying Channels," *submitted to IEEE Trans. on Signal Processing*, Jan. 2003. <http://acsp.ece.cornell.edu/pubJ.html>.
- [20] R. Negi and J. Cioffi, "Pilot tone selection for channel estimation in a mobile OFDM System ,", *IEEE Trans. on Consumer Electronics*, vol. 44, pp. 1122–1128, Aug. 1998.
- [21] S. Ohno and G. Giannakis, "Optimal training and redundant precoding for block transmissions with application to wireless ofdm ,", in *Proc. ICASSP*, (Salt Lake City, UT), May 2001.
- [22] S.Ohno and G.B.Giannakis, "Capacity maximizing pilots for wireless ofdm over rapidly fading channels ,", *submitted to IEEE Trans. Info. Theory*, April 2001.
- [23] C. Budianu and L. Tong, "Channel Estimation for Space-Time Block Coding Systems," *IEEE Trans. Signal Processing*, vol. 50, pp. 2515–2528, Oct 2002.
- [24] M. Dong and L. Tong, "Optimal Design and Placement of Pilot Symbols for Channel Estimation," *IEEE Trans. on Signal Processing*, vol. 50, pp. 3055–3069, December 2002.

- [25] R. G. Gallager, *Information Theory and Reliable Communication*. New York, NY: John Wiley and Sons, Inc., 1968.
- [26] R. Horn and C. R. Johnson, *Matrix Analysis*. New York, NY: Cambridge University Press, 1985.
- [27] T. Cover and J. Thomas, *Elements of Information Theory*. John Wiley & Sons, Inc., 1991.

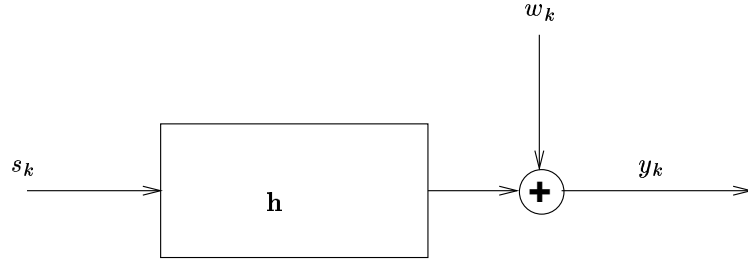


Fig. 1: System Model

figure

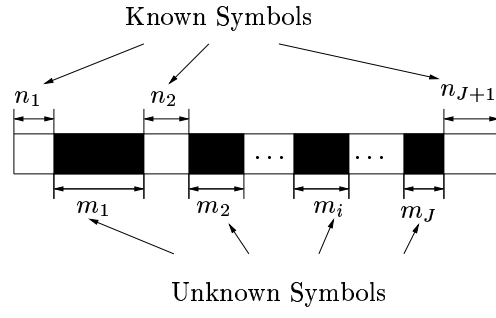


Fig. 2: Representation of Placement Schemes

figure

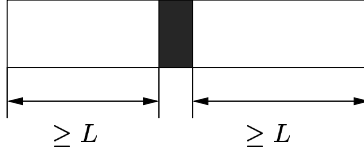


Fig. 3: A Sample Training Cluster

figure

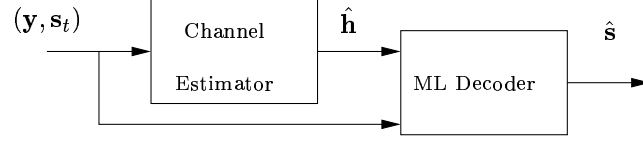


Fig. 4: Receiver Structure

figure

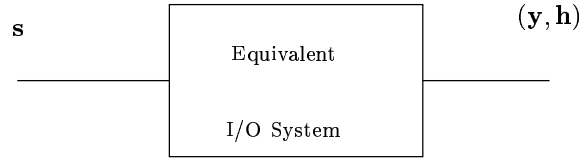
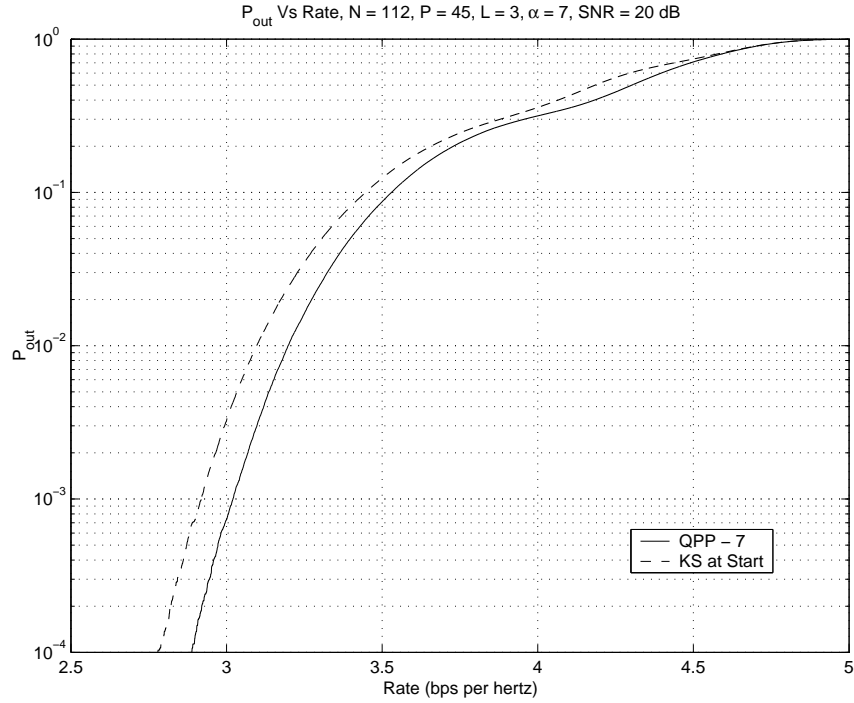


Fig. 5: Equivalent System

figure

Fig. 6: Performance of Placement Schemes at $\text{SNR} = 20\text{dB}$, $L=3$

figure

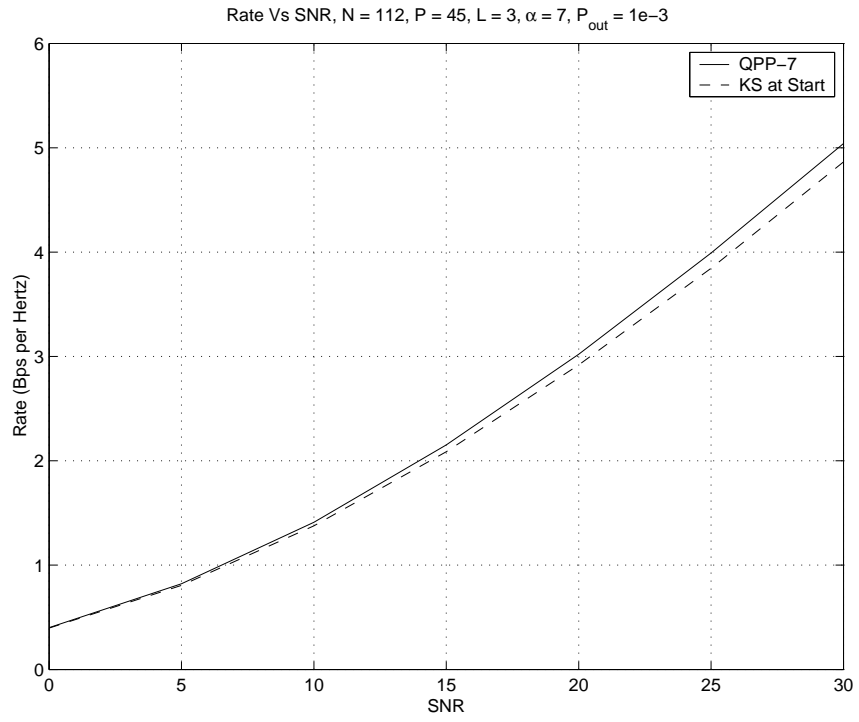


Fig. 7: Performance of Placement Schemes at $P_{out} = 10^{-3}$, $L = 3$

figure

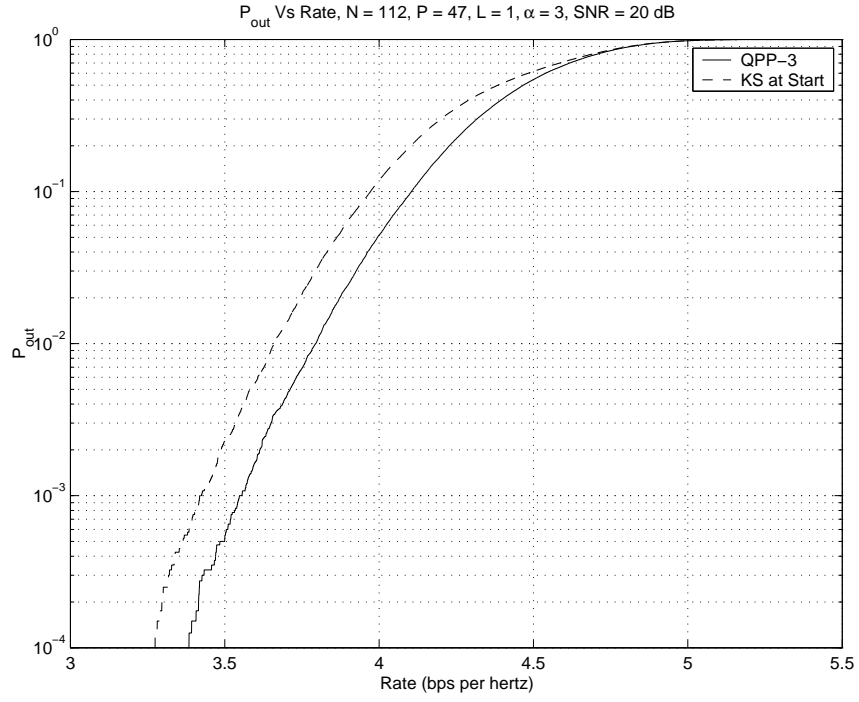


Fig. 8: Performance of Placement Schemes at SNR = 20dB, L=1

figure

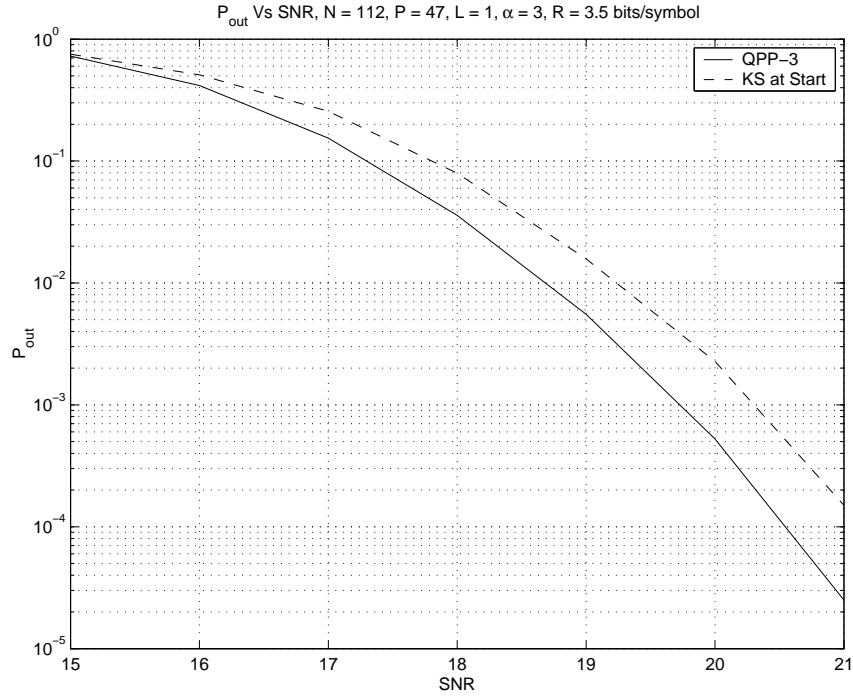


Fig. 9: Performance of Placement Schemes at R = 3.5 ,L = 1

figure

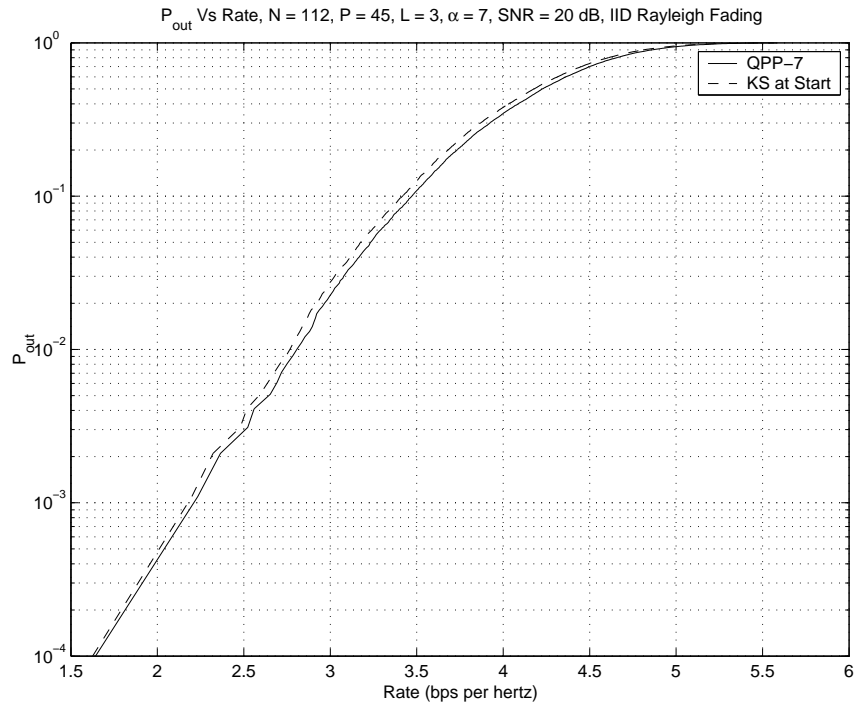


Fig. 10: Asymptotic Performance of Placement Schemes for Rayleigh Fading Channels SNR = 20dB ,L

figure = 3